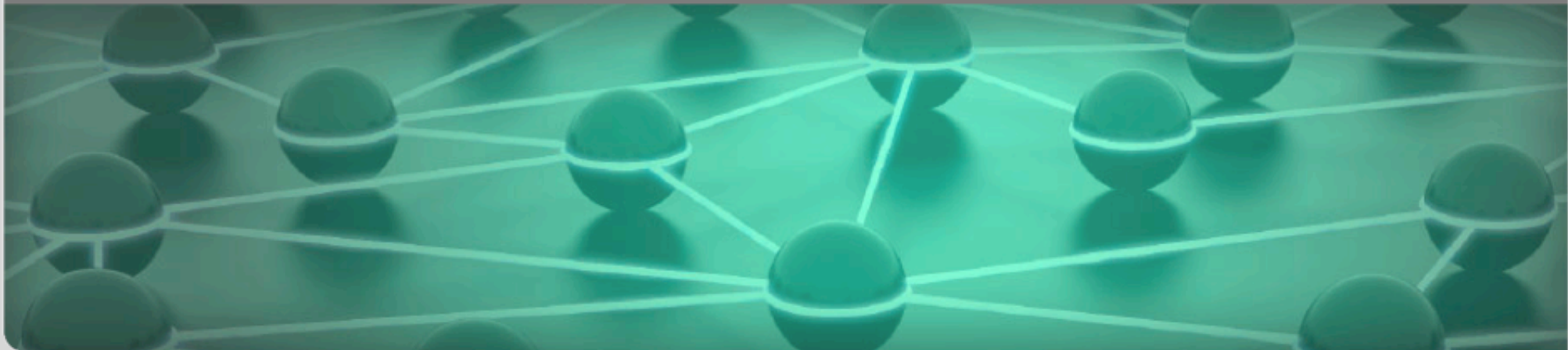


Logik im Semantic Web: Wunsch und Wirklichkeit

Sebastian Rudolph

28. AIK-Symposium "40 Jahre Angewandte Informatik Karlsruhe", 17.02.2012

Institut für Angewandte Informatik und Formale Beschreibungsverfahren



Prolog

AIFB

1997:



Institut für
Angewandte Informatik **AIFB**
und Formale Beschreibungsverfahren
Universität Karlsruhe (TH)

- 40 Jahre
- 143 PhDs, 20 Habils
- Mission:
Forschung und Lehre
Angewandte Informatik

Sebastian

1997:



- 35 Jahre
- 1 PhD, 1 Habil
- Mission:
Forschung und Lehre
Algebra & Logik



LS3: Wissensmanagement
& Semantic Web



Agenda

- Bestandsaufnahme
 - Wunsch
 - Wirklichkeit
- Ursachenanalyse
 - vier problematische Grundannahmen
 - mögliche Lösungsansätze

Die Idee des Semantic Web

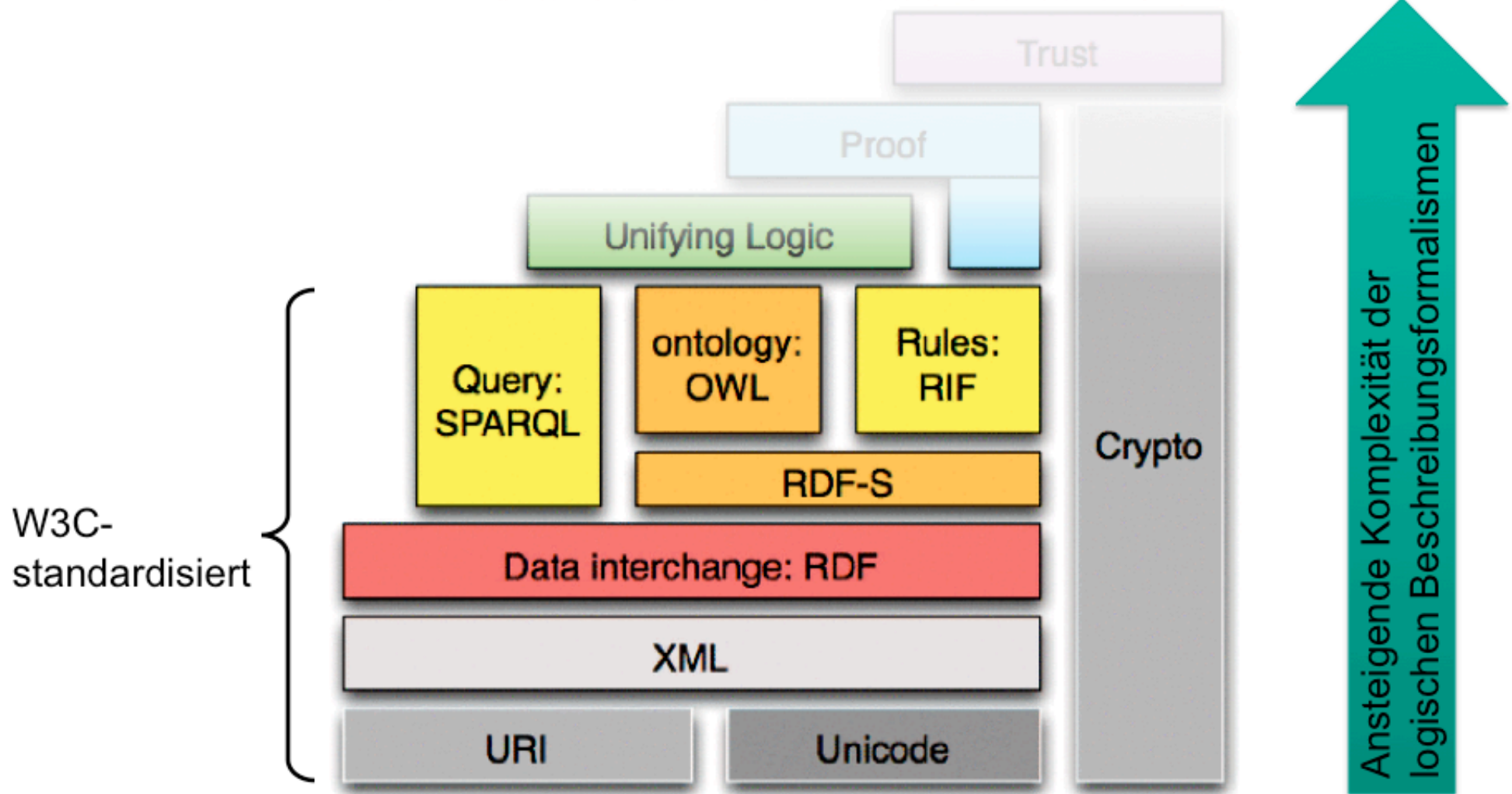


- Bereitstellung von Informationen/Wissen im Web in maschinenverarbeitbarer Form für intelligente Suche und Informationsintegration
- entscheidender Aspekt: *implizites Wissen* – Wissen welches nicht explizit gegeben ist, sondern sich aus der Kombination verschiedener Informationen herleiten lässt

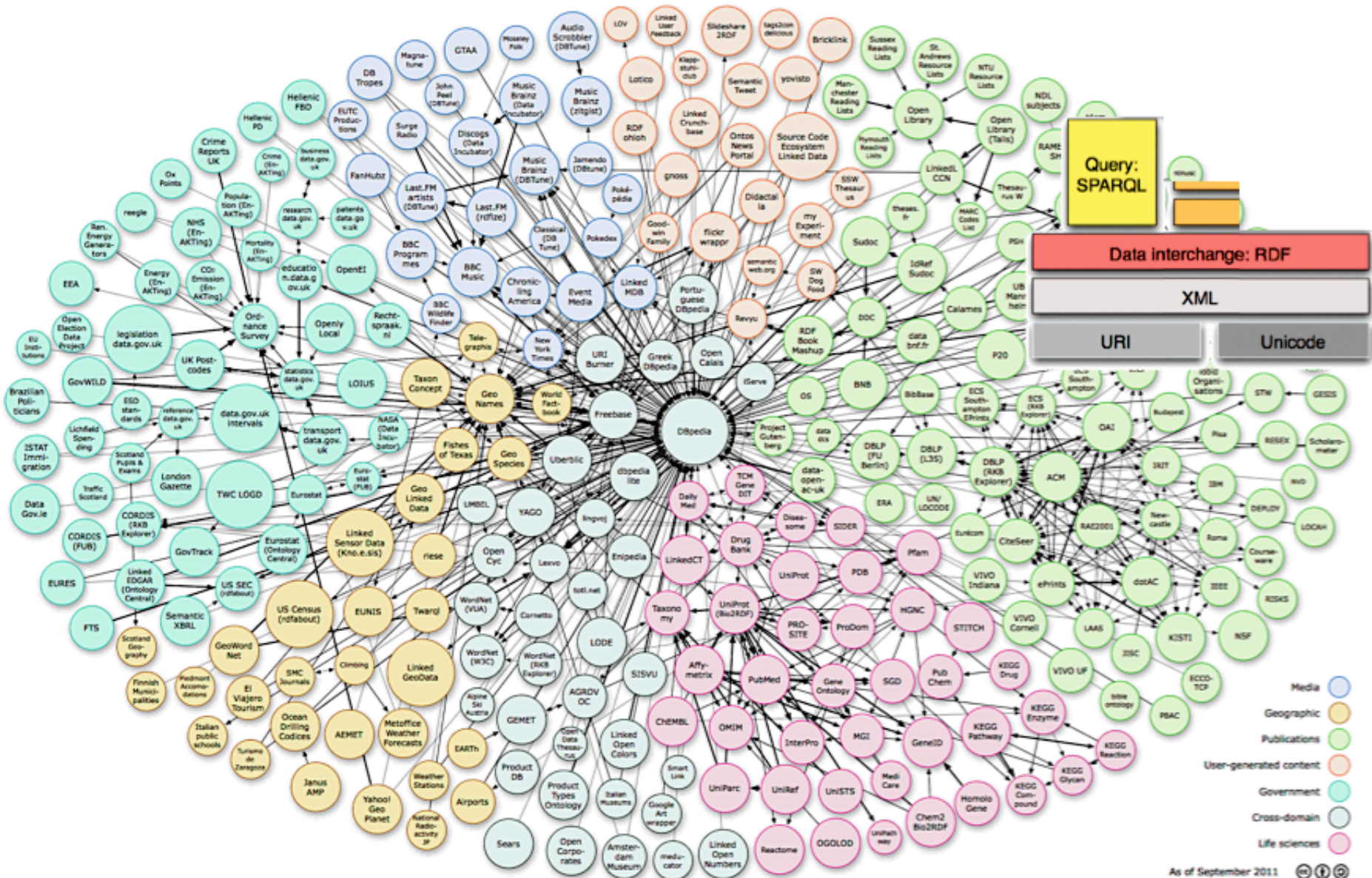
Voraussetzungen:

- offene Standards zur Beschreibung von Informationen mit klar definierter Semantik
- Inferenzmaschinen (sogenannte Reasoner) zur automatischen Bestimmung logischer Konsequenzen

Die Konzeption des Semantic Web à la W3C



Semantic Web = Linked Data Web?



Problematische Prämissen

- Entscheidbarkeit ist entscheidend.
- Menschen denken logisch.
- Wissen ist zweiwertig.
- Information ist quasistatisch.

1. Entscheidbarkeit ist entscheidend

„Entscheidbarkeit ist das primäre Kriterium für die praktische Verwendbarkeit eines logischen Formalismus.“



- grundlegendes Design-Prinzip bei der Definition von OWL:
maximale Ausdrucksstärke bei garantierter Entscheidbarkeit
- → Aufnahme vieler verschiedenartiger Modellierungsmittel
- → Komplizierte syntaktische Restriktionen

- andererseits: Entscheidbarkeit \neq effiziente Lösbarkeit
- Beispiel: logisches Folgern in OWL ist N2ExpTime-vollständig

Größe der Wissensbasis: n	Zeit für Anfragebeantwortung $2^{(2^n)}$
1	4
2	16
3	256
4	65.536
5	4.294.967.296
6	18.446.744.073.709.551.616



Lösungsansatz A: small is beautiful

Aufgabe der Ausdrucksstärke zu Gunsten von Einfachheit und Effizienzgarantie: tractable profiles

- Identifizierung „leichtgewichtiger“ Teilsprachen (tractable profiles) mit guten Berechnungseigenschaften hinsichtlich Laufzeit und Skalierbarkeit
- Verwendung von deduktiven Datenbanken oder Transformation der Anfragen nach SQL
- sinnvoll bei großen Datenmengen mit simplem Schema

Lösungsansatz B: no limits no guarantees

Aufgabe der Vollständigkeit zu Gunsten von Einfachheit und Ausdrucksstärke

- z.B. Prädikatenlogik erster Stufe als Modellierungssprache
- Verwendung von hochoptimierten Theorembeweisern (wie sie z.B. in der Hard- und Softwareverifikation eingesetzt werden), angepasst an große Datenmengen, neue Anfragearten und Unterstützung von anytime-Verhalten
- sinnvoll, wenn Vollständigkeit nicht essentiell ist

2. Menschen denken logisch

„Der durchschnittliche Semantic-Web-Anwender ist in der Lage, sein Wissen in eine logische Spezifikation zu überführen.“

- Laut kognitionspsychologischen Untersuchungen sind Menschen nur zu elementarsten logischen Schlüssen fähig:

modus ponens (aus „ $A \rightarrow B$ “ und „ A “ und folgt „ B “): 89-100%

modus tollens (aus „ $A \rightarrow B$ “ und „nicht B “ folgt „nicht A “): 25-57%

- Explizites logisches Modellieren ist noch anspruchsvoller als implizites Anwenden von logischen Schlussfolgerungen



2. Menschen denken logisch

- Noch schwieriger: sicherstellen, dass die spezifizierte Wissensmenge vollständig ist (d.h., dass jede relevante Anfrage eindeutig beantwortet werden kann).

Wissensbasis:

Es gibt Vögel.
 Alle Vögel legen Eier.
 Vögel sind keine Säugetiere.
 Eierleger sind nicht lebendgebärend.
 Es gibt lebendgebärende Säugetiere.

Es gibt eierlegende Säugetiere!

nicht beantwortbar:

Sind alle Säugetiere
 lebendgebärend?



Lösungsansatz: Intelligente User Interfaces

- Tools, mit deren Hilfe logisch nicht geschulte Nutzer intuitiv modellieren können
 - einfach bei simplen logischen Aussagen (Realisierung z.B. durch Tagging), schwierig bei komplexeren Sachverhalten
 - Unterstützung durch Techniken der maschinellen Sprachverarbeitung
 - flexible Aktualisierungsmechanismen

- Verfahren, die mit möglichst wenigen Fragen an Experten eine Wissensspezifikation vervollständigen
 - lokale Optimierung
 - formale Begriffsanalyse
 - induktive Logikprogrammierung

Intermezzo

Die „Flaschenhälse“ logischer Wissensrepräsentation:

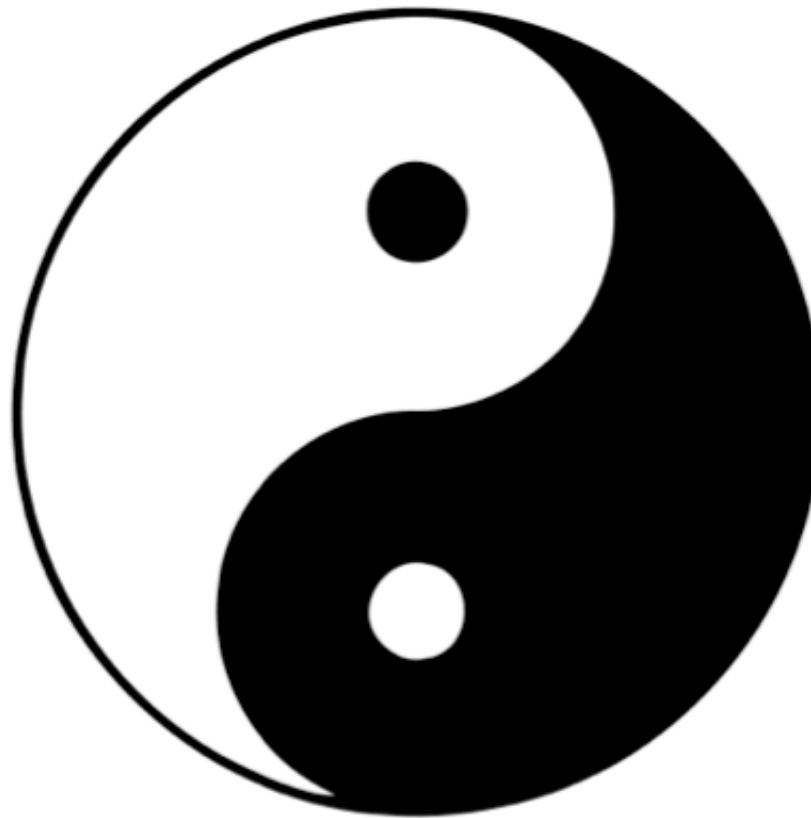
Modellierung

Anfragen



3. Wissen ist zweiwertig

*„Zweiwertige Logik ist die angemessene Art und Weise,
Wissensverarbeitung im Web zu betreiben.“*



Nicht wirklich: *ex falso (sequitur) quodlibet*

- aus einem Widerspruch lässt sich Beliebiges folgern; damit wird eine Wissensmenge durch einen einzigen „kleinen“ Widerspruch nutzlos im Sinne der ableitbaren Schlussfolgerungen

$$\frac{A \quad \neg A}{B}$$

- → klassische (zweiwertige) Logik ist nicht tolerant gegenüber Fehlern
- in typischen Web-Szenarien sind Fehler/Widersprüche aber kaum vermeidbar:
 - verschiedene Quellen / verschiedene Standpunkte
 - Verfahren der Informationsextraktion (nicht 100% präzise)

Lösungsansatz: tertium datur [sic]

- Nutzung von „robusteren“ fehlertoleranten Logiken, die mit unsicherem Wissen umgehen können
 - kontextualisierte Wahrheitswerte
 - fuzzy logic
 - parakonsistente Logiken
 - probabilistische Logiken (Markov logic)

- To do:
 - solide theoretische Grundlagen
 - Standardisierung

4. Information ist quasistatisch

„Information im Web wird persistent gespeichert und ändert sich nur sporadisch. Anfragen sind häufiger als Aktualisierungen.“



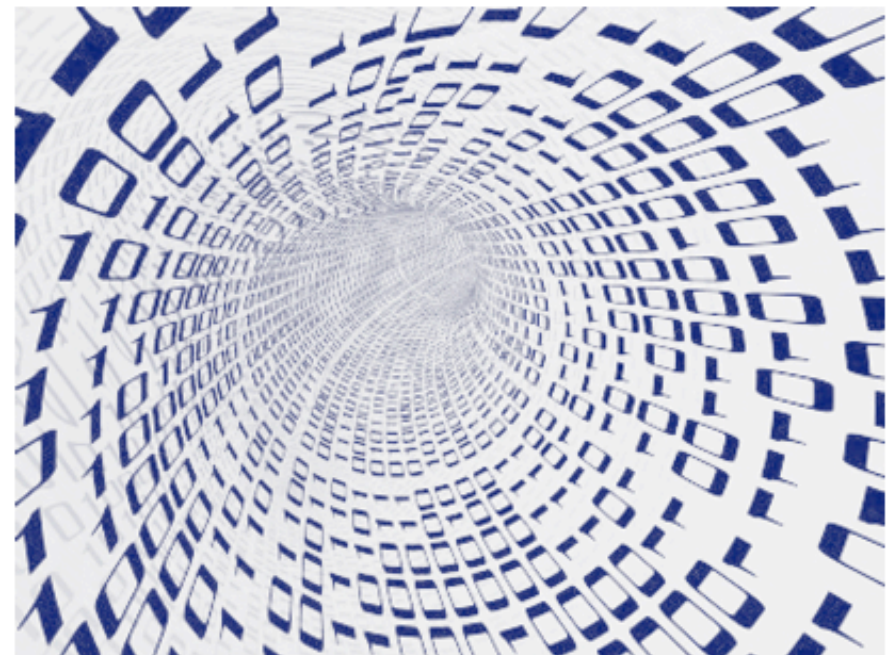
Weit gefehlt: πάντα ρεῖ

- News, Blogs, Tweets, Aktienkurse, Sensordaten
- Großteil der relevanten Web-Information ist hochgradig dynamisch
- Echtzeitverarbeitung ist essenziell
- herkömmliche Strategien zur Indexierung und zum automatischen Schlussfolgern sind ungeeignet



Lösungsansatz: stream reasoning

- Anleihen bei active und streaming databases und complex event processing
- Anreicherung um Inferenzalgorithmen die auf die Verarbeitung von Datenströmen optimiert sind
 - Logikprogrammierung
 - query rewriting
 - Parallelisierung
- To do:
 - solide theoretische Grundlagen
 - Standardisierung



Fazit & Credo

- formale Semantik zentral für Wissensrepräsentation (Herstellung von Interoperabilität)
- Logik ohne Scheuklappen:
 - mehr Flexibilität hinsichtlich Anforderungen
 - mehr Augenmerk auf Unterstützung der Modellierung
 - souveräner Umgang mit „verrauschter“ Information
 - zunehmend dynamischer Natur des Wissens Rechnung tragen

Vielen Dank